BACKGROUND INFORMATION

**Field/Industry:** Computer Science/Computer Vision and Machine Learning

**Summary of Work in Field/Industry:** I (Ehsan Fathi) right now work for JPMorgan Chase & Co. I am a Data Scientist working on developing models for fraud detection. Mainly using Computer vision and image processing techniques to solve problems like object detection, handwriting recognition, handwriting verification, signature verification, etc. Using the result of these in Machine learning models for different purposes.

I have worked for in this area through my Masters. For my thesis I worked on “A Scalable Solution for Extreme Multi-class Product Classification: An E-commerce Case Study”

I also worked on microscopic pathology pictures of lung cells to distinguish cancerous cells. (this is a process that take almost 10 years for humans to be trained for it.)

I have worked on machine learning algorithms and neural networks for about 10 years in different industries including healthcare, oil and gas, finance, marketing, etc.

**Summary of Top Achievements:**

- I have developed several algorithms and models for improving advertising, computer vision, object detection, text recognition, handwriting recognition, handwriting verification, signature verification.

- I have been granted Full-Scholarship for both undergrad and 2 Masters.

QUALIFICATIONS

* **Please indicate and describe the top 3-6 contributions you have brought to the field or industry using the template below.**

**1. “Check Fraud Detection”**

**Date of Project Initiation:** I started at July 2, 2018 but the project started a few months before I join the team.

**Date of Project Completion:** It’s an ongoing project

**Resulting Publication(s) (if applicable):** There is nothing published at the moment because it is all confidential and the corporate propriety but we are planning to publish a paper.

**Technical Summary of Work:**

Note: I have included several related and recent projects in this section. I can separate them and provide more information and details for each one of them. Please let me know if you think it is a better idea and is better to do it.

In order to solve the problem of detecting all the different object we couldn’t just use the usual object detection algorithms and models out there. The main problem is that object detection is actually combination of two problems: object localization and classification. The main difficulty in case of working on check images arises dealing with the later. Main reason is that all the classification models are designed to classify the common objects in real world as animals, flowers, human, car, furniture and so on that have a defined shape, but in case of objects in check images we are dealing with objects like: owner name, address, memo, signature, check number, etc. that do not have any specific shape. To clarify with an example, imagine the object at hand is a giraffe. We know a giraffe always appear with 4 legs, a long neck and a specific skin pattern. Now imagine a signature. How can we define a signature shape in general? There is no definition for signature shape because each one of them is unique. It can be a decent handwritten name or calligraphy of it or some lines with no meaning or even handwriting in other languages. This issue applies to other objects on the check. Like in a simple case both signature and payee name are handwritten names. How can we define a different shape for each one? Not possible. Imagine putting them next to each other and ask you which one is what. To solve this problem, I used a lot of ideas from different state of the art object detection models. Specially, YOLO. In the simplest way of explaining it has a backbone of Deep Neural Networks or more specifically Convolutional Neural Networks (Covnets). The output of the back bone network goes through a pipeline to determine the final result. It produces the final result as a list of objects that each one is defined by its name, the probability or the confidence score that model has in the final guess, coordination of the center of the object, width and height of it.

The result of this model is astonishing. I am not legally allowed to put any picture here but in case it is needed I can ask for the permission. This model has a very high performance which is comparable to all the state of the art general object detection models.

For signature verifier model, I used several ideas from face detection models. The big picture is that get goes through the history of the check images for the costumer at hand and gets several of the genuine signatures and runs them through a Convolutional Neural Network to crunch it to a vector. We do the same to the current signature and then feed these results to and XGBoost layer.

Ideas for handwriting recognition and authentication are considered highly confidential at them moment. But if it is needed I can ask for permission to provide details.

**Plain Language Summary of Work:**

Note: I have included several related and recent projects in this section. I can separate them and provide more information and details for each one of them. Please let me know if you think it is a better idea and is better to do it.

Right now I am working on a revolutionary project for banking fraud detection. At the moment of writing this document JPMorgan Chase and also all the other banks and their costumers are loosing a huge amount of money every day because of fraud in all different areas like stolen/copied credit cards, fraudulent checks, etc. (Numbers are in million dollars in month. I can provide numbers if needed). The reason is that banks do not have a technology that can do tasks like check recognition automatically with high accuracy in real-time. By check recognition I mean finding different objects in a check like: signature, name, address, payee name, amount, amount in words, memo, check number, MICR numbers including routing number and account number and so on. To the best of my knowledge no bank in the united states has a model in work that can find these objects in the check image. This is a very difficult problem to solve due to several reasons. First of all, there are countless number of check designs even in one bank. There are customized check designs for different seasons, holidays, or personal preferences. Also considering printed checks each company has its own design and template. This problem keeps banks from being able to proceed in utilizing the information in the check image.

You may have noticed that you can sign your check very carelessly and it goes through, some people have different signatures and all of them go through! The reason is that even if we have can recognize the signature there is no model for signature verification in work in banks. Same problem for handwriting verification and authentication. Models that are in production right now for fraud detection are very basic models and not accurate.

I am the lead researcher of a R&D (Research and Development) group to solve these problems and solve finance technology problems using modern approaches in Artificial Intelligence like Deep Neural Networks or Deep learning. To solve the above mentioned problem I had a very important role and in a short amount of time I had a great progress. To the best of my knowledge we are the first group who have done this. I developed a high accuracy real-time object detection model to detect all the different objects on the check as mentioned above. It can also classify each object as handwritten or printed text which is a very important feature. Because later these parts need to be parsed and at the time of parsing it is very important to know if the detected text is printed or handwritten. Because they have completely different mechanism and approaches.

After this achievement I proceeded in making a model to verify the signature. To the best of my knowledge it is the first industrial level signature verification model based on modern Artificial intelligence approaches. It can verify the authenticity of the signature by 98% accuracy.

Next we parsed the printed texts. I am part of the team who did it and this part is not done just by myself.

Right now my model can catch a huge amount of the fraud which is not comparable to the models in production (in use) right now. The first phase of it will be implemented in 2019. Right now I am working on parsing the handwritten parts of the check for two different purposes. First, the obvious reason is to convert the picture to text so we can use the information. Second, for handwriting authentication. This is a very difficult problem to solve in general. In the case of check images, it is even more difficult. Because usually there is a huge overlap between payee name and amount in words which is the line under payee. Also, check images have low quality which makes it more difficult.

Please do not hesitate to let me know if any part is not clear of more detail is needed.

**Summary of the Significance of the Work:**

The reason these series of projects are important and significant is that all of them have been developed for the first time in the world. As far as I know no other comparable example exists. It also has a huge impact on both FinTech (Finance Technology) industry, computer science community and nation. The impact on the FinTech industry comes from the fact that it prevents a huge amount of check fraud. So both banks, credit companies and people of the nation will have benefits from it. Handwriting Recognition and Handwriting Verification and Authentication are the open research areas of research and my works contribute to computer science community.

**Summary of the Implementation/Influence of the Work:**

All that I did in these series of projects except handwriting recognition and handwriting verifier is going through production in 2019. It is a benchmark because it is going to affect “inclearing” checks which means checks that Chase costumers write to Chase costumers. Next phase is going to expand it to all the incoming checks from different banks and other sources. All the models that have been developed are general purpose models. For example, signature verifier can be used to verify, validate and authenticate any signature on any document. The progress in my work was so outstanding that is not comparable to products and models that vendors are providing. Because of such an outstanding progress in such a short time it already has been presented to Jamie Dimon, JPMorgan Chase’s CEO, and has been considered to invest more on it. I can provide evidences for it.

**“My graduate studies”**

**Date of Project Initiation: Jan 2016 – May 2018**

**Date of Project Completion: May 2018**

**Resulting Publication(s) (if applicable): 2 book chapters**

[Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning](javascript:void(0))

[Deep Neural Networks for Natural Language Processing](javascript:void(0))

**Technical Summary of Work:**

I designed a scalable and flexible architecture of different size of convolutional neural networks that is practical in real world cases and is practical for huge dataset which are very unbalanced. The beauty of the architecture is that as the system grows and dynamics of the system may rule out some of the products, the efficiency of the architecture stay consistent. In other words, it scales properly without a big maintenance cost. It combines the divide and conquer principles and complex system analysis techniques with Convolutional neural networks and modern classification techniques to achieve this goal. In the benchmark provided as the case study in my thesis just uses well known architectures as lego blocks of the architecture to make the comparison easy and also make it possible to reproduce the results easily. Any other networks can be used as well. The result shows that using this architecture this task can be accomplished in on tenth of the time with much less resources in cost of reasonable loss of accuracy. Top-1 score metric has been used for comparison between different cases. Using other metrics like Precision and Recall we observe that not only we don’t loose any accuracy but also the proposed architecture is able to improve it.

**Plain Language Summary of Work:**

I worked in the area of Machine learning and Artificial intelligence during my graduate studies. More specifically on Image Processing and Natural Language Processing. I worked on different projects. One of the more interesting ones was distinguishing the cancerous cells in microscopic pathology pictures of lung cells. I worked on the classification problem in the real world size. Image classification is the main task in image processing. Although, there were a lot of advances in recent years, it is still quite a challenge. On the other hand, due to the progress in technology, e-commerce has emerged as the fastest-growing sector of the U.S. marketplace. Product classification is an extremely important issue in e-commerce. In this work, we propose a scalable, flexible, practical, modular and efficient architecture to use image classification techniques for product classification just using product images. Considering the diversity of products offering in retail online retail stores it is not surprising that we confront an excessive number of classes. Case study is Cdiscount which is the biggest non-food e-commerce company in France which has made about 3 billion euros in 2017. They have about 30 million products up for sale. We focused on developing a CNN architecture to tackle this challenge and provide a more general, flexible, scalable and efficient solution for Cdiscount image classification business problem. Results of applying the proposed architecture shows a reasonable accuracy which shows the efficiency of the architecture.

**Summary of the Significance of the Work:**

(*Please provide one paragraph describing the general significance of this work. Why is work in this area important to the field and the nation?*)

**Summary of the Implementation/Influence of the Work:**

(*Please provide at least one paragraph describing how this work has already been influential in the field. You should provide specific examples of where and how your work has been used by others, giving as much detail as possible. Examples of implementation include (but are not limited to) licenses or contracts, collaborations, technology transfer agreements, patent commercialization, clinical guidelines, clinical trials, or emails asking for your work.*)

**“TITLE OF PROJECT”**

**Date of Project Initiation:**

**Date of Project Completion:**

**Resulting Publication(s) (if applicable):**

**Technical Summary of Work:**

(*Please provide two paragraphs describing the work using the technical terminology and explanation of work that an expert would be able to understand and evaluate. The summary should be around 300 words in length.)*

**Plain Language Summary of Work:**

(*Please provide two paragraphs describing the work in layman’s terms or plain language that an average person without special expertise would be able to understand and evaluate. The summary should be around 300 words in length.*)

**Summary of the Significance of the Work:**

(*Please provide one paragraph describing the general significance of this work. Why is work in this area important to the field and the nation?*)

**Summary of the Implementation/Influence of the Work:**

(*Please provide at least one paragraph describing how this work has already been influential in the field. You should provide specific examples of where and how your work has been used by others, giving as much detail as possible. Examples of implementation include (but are not limited to) licenses or contracts, collaborations, technology transfer agreements, patent commercialization, clinical guidelines, clinical trials, or emails asking for your work.*)

**“TITLE OF PROJECT”**

**Date of Project Initiation:**

**Date of Project Completion:**

**Resulting Publication(s) (if applicable):**

**Technical Summary of Work:**

(*Please provide two paragraphs describing the work using the technical terminology and explanation of work that an expert would be able to understand and evaluate. The summary should be around 300 words in length.)*

**Plain Language Summary of Work:**

(*Please provide two paragraphs describing the work in layman’s terms or plain language that an average person without special expertise would be able to understand and evaluate. The summary should be around 300 words in length.*)

**Summary of the Significance of the Work:**

(*Please provide one paragraph describing the general significance of this work. Why is work in this area important to the field and the nation?*)

**Summary of the Implementation/Influence of the Work:**

(*Please provide at least one paragraph describing how this work has already been influential in the field. You should provide specific examples of where and how your work has been used by others, giving as much detail as possible. Examples of implementation include (but are not limited to) licenses or contracts, collaborations, technology transfer agreements, patent commercialization, clinical guidelines, clinical trials, or emails asking for your work.*)

* **Please list and describe any nationally or internationally-recognized prizes or awards for excellence you have received in the field/industry.**

(Note: *Please fill out the following information for each award that you have won. You should also upload evidence of this information. If the award was given to a sponsoring institution or supervisor, please provide the name of the person who can write a letter attesting to your contributions on the award-winning project.*)

**1. “TITLE OF AWARD”**

**Award Recipient:**

**Awarding Institution:**

**Who is Eligible to Compete:**

**Number of Competitors/Winners:**

**Selection Criteria:**

**Who are the Judges:**

**Notable Past Award Winners:**

* **Please list and describe any memberships or professional organizations of which you are a member.**

(Note: *Please fill out the following information for each organization of which you are a member. You should also upload a certificate or letter as evidence of your membership, and provide the relevant sections of the organization’s bylaws or constitution which discuss the requirements for membership.*)

**1. “NAME OF ORGANIZATION”**

**Level of Membership:**

**Requirements for Membership:**

**Who Judges Membership Eligibility:**

* **Please indicate any media coverage you have received.**

(Note: *Please fill out the following information for each piece of media coverage. You should also upload or provide a link to the actual article*.)

**1. “TITLE OF ARTICLE”**

**Date Published:**

**Author:**

**Name of Journal/Magazine/Newspaper/Website:**

**Circulation/Number of People who Read the Publication:**

**Summary of Article’s Focus:**

**Relevance to Your Original Work:**

* **Please indicate and describe any occasions where you have served as a judge of the work of other experts in the field/industry.**

(Note: *Please fill out the following information for each instance of judging that you have actually completed. Invitations to judge are not sufficient. You should also upload evidence that you actually completed this work, which can include thank you emails. If you are not able to provide thank you emails or similar evidence, we can draft a letter for the organization to sign accordingly*)

**1. Organization (Journal, Conference, etc.) for which you Judged:**

**Number of reviews completed:**

* **Please provide a complete list of articles, books, chapters, conference proceedings, workshops, or seminars where you have published or presented.**

(Note: *Please include the title of article/presentation, name of publication, date, credited authors, and any relevant publication information. If this list with the related information is readily available on your resume or curriculum vitae, you may copy and paste this information or note such below*.)

**1.**

* **Please describe any leading or critical roles you have played in prominent organizations in the field/industry.**

(Note: *Please provide as much detail about your roles and responsibilities as you can so that we may discuss the significance and importance of these contributions to your employing organization)*

**1. “TITLE OF POSITION”**

**Name of Organization:**

**Dates of Employment:**

**Brief Summary of Organization’s Prestige/Reputation in Field:**

**Brief Summary of Role and Responsibilities:**

**Summary of Key Contributions to Organization:**

* **Please indicate your annual salary or equivalent remuneration for services:**

Three-Prong Test (Matter of Dhanasar)

**1a. How is your field or industry inherently beneficial to the interests of the United States (i.e., how is work in your field of “substantial merit”)?**

**1b. How does your work in particular benefit the United States (i.e., how is your work of “national importance”)?**

**2. How do your past accomplishments position you to continue contributing to your field? What kind of major contributions to your field have you made in the past?**

**3. Is there anything else you can tell us that really sets you apart from you peers? This can include specific particularly impressive contributions that have had a large impact on your field, awards or other recognition you have received, or particular skills that you have that go beyond what would be expected of an average individual in your field.**

* **Please include any additional information below that you feel speaks to the substantial importance or the distinguished merit of your past, current, and future professional endeavors:**